The goal of this problem is to show that there are two equivalent expressions for the residual sum of squares in linear regression.
Let our training data be $\{(\mathbf{x}_1, y_1), \ldots, (\mathbf{x}_N, y_N)\}$, where the vector $\mathbf{x_n} \in \mathbb{R}^D$ includes the D features and $y_n \in \mathbb{R}$ is the label of sample n.

The training data can be also written in a matrix/vector notation as:

$$\mathbf{X} = \begin{bmatrix} 1 & \mathbf{x}_1^T \\ & \vdots \\ 1 & \mathbf{x}_N^T \end{bmatrix} = \begin{bmatrix} 1 & x_{11} & x_{12} & \ldots & x_{1D} \\ & & \vdots & & \\ 1 & x_{N1} & x_{12} & \ldots & x_{ND} \end{bmatrix} \in \mathbb{R}^{N \times (D+1)} \text{ and } \mathbf{y} = [y_1, \ldots, y_N]^T \in \mathbb{R}^N$$

where x_{nd} is the d^{th} feature of sample n.

Let's also assume that the weight of the linear regression model is written as $\mathbf{w} = [w_0,\ w_1,\ \ldots,\ w_D]$, where w_0 is the bias.

Show that the following expressions of the RSS error are equivalent:

$$RSS(\mathbf{w}) = \sum_{n=1}^{N} \left[y_n - \left(w_0 + \sum_{d=1}^{D} w_d x_{nd} \right) \right]^2$$

$$RSS(\mathbf{w}) = (\mathbf{y} - \mathbf{X}\mathbf{w})^T (\mathbf{y} - \mathbf{X}\mathbf{w})$$

$$\vec{y} - \underline{X}\vec{w} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_N \end{bmatrix} - \begin{bmatrix} 1 & x_{11} & x_{12} & \ldots & x_{1D} \\ 1 & x_{21} & x_{22} & \ldots & x_{2D} \\ \vdots & & & & \\ 1 & x_{N1} & x_{N2} & \ldots & x_{ND} \end{bmatrix} \begin{bmatrix} w_0 \\ w_1 \\ w_2 \\ \vdots \\ w_D \end{bmatrix}$$

→ the feature of each sample is multiplied by the corresponding weight.

$$= \begin{bmatrix} y_1 - (w_0 + w_1 x_{11} + w_2 x_{12} + \ldots + w_D x_{1D}) \\ \vdots \\ y_N - (w_0 + \underbrace{w_1 x_{N1} + w_2 x_{N2} + \ldots + w_D x_{ND}}) \end{bmatrix} \in \mathbb{R}^N$$

→ $\sum_{d=1}^{D} w_d x_{Nd}$

$$\underbrace{(\vec{y} - X\vec{w})^T}_{\mathbb{R}^{1\times N}} \underbrace{(y - X\vec{w})}_{\mathbb{R}^{N\times 1}} = \left[y_1 - \left(w_0 + \sum_{d=1}^{D} w_d x_{1d}\right), \ldots, y_N - \left(w_0 + \sum_{d=1}^{D} w_d x_{Nd}\right) \right] \begin{bmatrix} y_1 - \left(w_0 + \sum_{d=1}^{D} w_d x_{1d}\right) \\ \vdots \\ y_N - \left(w_0 + \sum_{d=1}^{D} w_d x_{Nd}\right) \end{bmatrix}$$

$$= \left(y_1 - \left(w_0 + \sum_{d=1}^{D} w_d x_{1d}\right)\right)^2 + \ldots + \left(y_N - \left(w_0 + \sum_{d=1}^{D} w_d x_{Nd}\right)\right)^2$$

$$= \sum_{n=1}^{N} \left[y_n - \left(w_0 + \sum_{d=1}^{D} w_d x_{nd} \right) \right]^2$$

Let $\mathbf{x} = [x_1, x_2] \in \mathbb{R}^2$ be a 2-dimensional vector.

(1) Compute the first derivative and Hessian of the function f.

$$\frac{\partial f(\vec{x})}{\partial \vec{x}} = \left[\frac{\partial f(\vec{x})}{\partial x_1} \quad \frac{\partial f(\vec{x})}{\partial x_2} \right]^T = \left[2(x_1 - 3x_2) \quad -6(x_1 - 3x_2) \right]^T \in \mathbb{R}^2$$

$$H_{f(\vec{x})} = \begin{bmatrix} \frac{\partial^2 f(\vec{x})}{\partial x_1^2} & \frac{\partial^2 f(\vec{x})}{\partial x_1 x_2} \\ \frac{\partial f(x)}{\partial x_1 \partial x_2} & \frac{\partial^2 f(\vec{x})}{\partial x_2^2} \end{bmatrix} = \begin{bmatrix} 2 & -6 \\ -6 & 18 \end{bmatrix} \in \mathbb{R}^{2\times 2}$$

(2) Show that the function $f(\mathbf{x}) = (x_1 - 3x_2)^2$ is convex by showing that its Hessian is positive semi-definite.

For any $\vec{u} \in \mathbb{R}^2$, $\vec{u} = [u_1, u_2]^T$, we have:

$$\vec{u}^T H_{f(\vec{x})} \vec{u} = [u_1 \; u_2]^T \begin{bmatrix} 2 & -6 \\ -6 & 18 \end{bmatrix} \begin{bmatrix} u_1 \\ u_2 \end{bmatrix}$$

$$= [2u_1 - 6u_2, -6u_1 + 18u_2]^T \begin{bmatrix} u_1 \\ u_2 \end{bmatrix}$$

$$= 2u_1^2 - 6u_1u_2 - 6u_1u_2 + 18u_2^2$$

$$= 2u_1^2 - 12u_1u_2 + 18u_2^2$$

$$= 2(u_1^2 - 6u_1u_2 + 9u_2^2) = 2(u_1 - 3u_2)^2 \geq 0$$

Therefore $H_{f(\vec{x})}$ positive semi-de

So $f(\vec{x})$ is conve

(3) Show that the function $f(\mathbf{x}) = (x_1 - 3x_2)^2$ is convex by showing that the eigenvalues of its Hessian are non-negative.

$$\operatorname{Det}(H_{f(\vec{x})}) = 0 \Rightarrow \begin{vmatrix} \lambda - 2 & -6 \\ -6 & \lambda - 18 \end{vmatrix} = 0 \Rightarrow \lambda^2 - 20\lambda + 36 - 36 = 0 \Rightarrow$$

$$\Rightarrow \lambda(\lambda - 20) = 0 \Rightarrow \lambda = 0 \text{ and } \lambda_2 = 20$$

Both eigenvalues are non-negative, therefore $H_{f(\vec{x})}$ is positive semi-defite, so f is convex.